



Lagring av forskningsdata i Norge

Hva tilbyr de største arkivene?

Therese Skagen

Handelshøyskolen BI

2015

Prosjektet er støttet av Nasjonalbiblioteket, prosjektnummer 2013/397.

Materialet i denne publikasjonen kan brukes fritt på den betingelse at forfatter krediteres og full referanse oppgis.

Skagen, Therese (2015). Lagring av forskningsdata i Norge – Hva tilbyr de største arkivene?
Prosjektrapport. Oslo: BI-biblioteket. Handelshøyskolen BI. Tilgjengelig fra: <http://brage.bibsys.no/>

Forord

Handelshøyskolen BI har i løpet av de siste årene arbeidet med å lage sitt eget lokale, institusjonelle arkiv for forskningsdata. Prosjektet BIRD er et pilotprosjekt for å lage en struktur som forskere ved Handelshøyskolen BI kan benytte seg av for å lagre og dele deres forskningsdata. Prosjektet befinner seg i en utviklingsfase til pilotversjon. BIRD er et lokalt arkiv som utvikles i samarbeid med BIBSYS. Arkivet bygger videre på åpen kildekode programvare brukt ved institusjonene for å danne publikasjonsarkiver, Dspace. Det foregår en utvikling i retning av å tilpasse arkivene for forskningsdata. Utviklingen legger vekt på å følge internasjonale standarder som kreves av EUs OpenAire satsing, med anbefalinger omkring bruk av DataCite kriterier for registreringspraksis og utveksling av metadata. Dette er i tråd med dagens praksis hvor institusjonelle arkiver anvender i dag metadatastanden Dublin Core. Standarder for utveksling av metadata for å øke gjenbruk av data gjøres gjennom OAI-PMH protokollen, slik som dagens bruk i publiseringsarkivene.

Prosjektet ble våren 2014 tildelt midler fra Nasjonalbiblioteket for å ta stilling til hvordan prosjektet BIRD forholder seg til andre aktører knyttet til levering og lagring av forskningsdata. Det lokale arkivet BIRD sin kobling til andre aktører er aktuell for å se på hvordan aktørene i sektorene kan utfylle hverandre.

Det er av interesse å se på hvilke tilbud eller muligheter som for øyeblikket befinner seg i en norsk kontekst. Det finnes i dag en rekke løsninger for lagring av forskningsdata i Norge og internasjonalt som forskere ved høyere utdanningsinstitusjoner benytter seg av. Det kan sees en begynnende avklaring omkring hva som kan være aktuelle arbeidsoppgaver og arbeidsfordeling mellom utdanningsinstitusjonen og dataarkiv. Dette er av interesse for å kunne videreformidle hvilket tilbud som finnes til forskerne ved institusjonene.

En kartlegging av norske arkiver vil bidra til å vise hvilken retning eventuelt BIRD prosjektet kan gå, og mulige koblinger til andre aktører. På denne måten kan en kartlegging bidra til å avklare hvilke roller biblioteksektoren / BIBSYS kan ivareta i samarbeid med fagbibliotekene og hvor grenseflatene bør gå mot øvrige aktører.

Oslo, april 2015.

Sammendrag

I løpet av de to siste årene er det kommet krav fra forskningsfinansiører omkring tilgjengeliggjøring av forskningsdata når prosjekter har fått økonomisk støtte (European Commission, 2013; Norges forskningsråd, 2014b). Kravene innebærer at forskeren må vurdere å gjøre informasjon om digitale forskningsdata tilgjengelig og lagre data over lengre tid. Kravene omtaler digitale resultatdata, som lagres og bevares over tid slik at det kan gjenbrukes. Kravene innebærer at forskningsdataarkiver som infrastruktur må fungere på en bestemt måte. Lagring og bevaring er en avgrenset del av forskningsdatas livssyklus (Jones, Pryor, & Whyte, 2013).

Kravene gjør det interessant å se på hvilke tilbud og funksjonalitet som finnes for avlevering, lagring og deling av forskningsdata. Internasjonalt er det identifisert funksjonalitet og støttetjenester som kan benyttes i arkiver for forskningsdata (Dillo & De Leeuw, 2014; PARADE, 2009; Reilly, Schaller, Schrimpf, Smit, & Wilkinson, 2011). Samlet utgjør funksjonaliteten og støttetjenester kriterier på en kvalitetsmessig standard som dataarkiver vurderes etter før man sier at de tilbyr gode tjenester for lagring og bevaring av forskningsdata.

Gjennom en kartlegging av aktiviteter blant norske aktører for lagring av forskningsdata vil det fokuseres på følgende forskningsspørsmål: Hvilket tilbud gir Norges største aktører for avlevering og lagring av forskningsdata som forskere knyttet til høyere utdanning og forskningsinstitutter kan benytte seg av? Hvilke planer har aktørene for videre arbeid? Denne casestudien ser på de største, nasjonale aktørene for datalagrings tilbud til forskerne i Norge, arkivet til Norsk samfunnsvitenskapelige datatjeneste og NorStore Research Data Archive.

Funksjonalitetene i arkivene er lignende, slik at dette er noe overlappende mellom de to arkivene. Forskjeller i formater og størrelse på data som arkivene kan motta og bevare over tid gjør arkivene forskjellige, slik at de utfyller hverandre. Dersom man tar høyde for utviklingen av forskningsdatainfrastrukturen i Norge som er planlagt, vil de nasjonale tjenestene fra dataarkivene gi de største brukergruppene av forskerne et tilbud for bevaring av data ved bruk av de mest vanlige filformatene.

Innhold

Forord	2
Sammendrag	3
Innledning - Lagring av forskningsdata i Norge	5
Problemstilling.....	5
Metode	6
Lagring og bevaring av forskningsdata	7
Norske aktører for lagring av forskningsdata	8
Hvilken funksjonalitet bør arkivet ha?	9
Kartleggingskriterier/ Intervjuguide	11
Beskrivelser av norske arkiver	15
NorStore Research Data Archive	15
Norsk Samfunnsvitenskapelig Datatjeneste – Arkivet	18
Likhet og ulikheter i tjenester for lagring og bevaring av forskningsdata	21
Avrunding	23
Etterord	25
Referanser	30

Innledning - Lagring av forskningsdata i Norge

Problemstilling

Både nasjonalt og internasjonalt er det en pågående diskusjon om hvilken rolle fagbiblioteket skal ha i arbeidet med å sikre en fremtidsrettet infrastruktur for forvaltning og deling av forskningsdata.

Politiske myndigheter har signalisert ønske om større grad av åpenhet og krav til deling av forskningsdata. I Forskningsmeldingen kap. 3.4.3. sies det at "Bedre tilgang til forskningsdata bidrar til å forenkle forskningssamarbeid og til å høyne kvaliteten på forskningen. Regjeringen ønsker å legge til rette for økt tilgjengeliggjøring av offentlige forskningsdata." (Kunnskapsdepartementet, 2013).

Norges forskningsråd [NFR] lanserte i september 2014 sine første retningslinjer for arkivering og deling av forskningsdata. Retningslinjene vektlegger at forskningsdata innhentet over midler fra NFR skal stilles tilgjengelig for andre forskere. Dette er en oppfølging av Forskningsrådets nasjonale strategi for forskningsinfrastruktur og et nytt veikart over strategisk viktige infrastrukturer (2012), og i henhold til de signaler som er gitt av EU i Horizon 2020 (European Commission, 2013).

Bibliotekorganisasjonene Research Information Network [RIN], US Council on Library and Information Resources [CLIR], The Association of European Research Libraries [LIBER] har foretatt undersøkelser som viser at bibliotekarer ser oppgaver knyttet til forskningsdata som en naturlig utvidelse av deres roller etter hvert som forskningspubliseringsprosessen endrer seg (Reilly et al., 2011; Thestrup et al., 2012; Van der Graaf & Waaijers, 2011).

Bibliotek og datasenter er enig om at data bør bevares, spesielt om forskningsprosjekter er offentlig finansiert, men det er få av bibliotekene og senterne som tilbyr denne typen tjenester (Kuipers, Hoeven, & Insight, 2009). Arbeid med lagring og gjenbruk av forskningsdata er et omfattende felt. Det er omdiskutert i litteraturen om hvilken rolle bibliotekene skal ha, og hvordan arbeid kan fordeles på forskjellige aktører og hvordan de kan samarbeide (Borgman, 2015; Nielsen, 2012). Likeledes diskuteres det hvilke tjenester som skal finnes lokalt ved en utdanningsinstitusjon/arbeidsplass for forskere, hvilke som kan kobles til fagmiljø fortrinnsvis internasjonalt og hvilke tjenester som bør være tilgjengelige på et nasjonalt plan (Kruse & Thestrup, 2014; Reilly et al., 2011). Kunnskap om tilbud og planer til eksisterende forskningsdataarkiver i Norge kan være å bidra inn i diskusjoner omkring oppgaver og arbeidsfordeling mellom forskningsinstitusjonene, fag- og forskningsbibliotekene og dataarkiv. Grunnlaget kan være et moment om bibliotekenes eventuelle roller i forhold til forskningsdata.

Gjennom en kartlegging av aktiviteter blant norske aktører for lagring av forskningsdata vil det fokuseres på følgende forskningsspørsmål:

- Hvilket tilbud gir Norges største aktører for avlevering og lagring av forskningsdata som forskere knyttet til høyere utdanning og forskningsinstitutter kan benytte seg av?
- Hvilke planer har aktørene for videre arbeid?

Metode

Denne caseundersøkelsen er sammensatt av flere faser. Den første fasen omhandlet fastsetting av kriterier som arkivene kan vurderes etter. Disse kartleggingskriteriene ble utarbeidet basert på en gjennomgang av aktuell litteratur angående tekniske spesifikasjoner omkring lagring og deling av data, målgruppe og tilbud om støttetjenester. Utarbeidelse av kartleggingskriterier ble basert på litteratur fra

- a. PARADE White Paper (2009) som henviser til en rekke tekniske karakteristika for vitenskapelig e-infrastruktur bør inkluderes.
- b. Data Seal of Approval (Dillo & De Leeuw, 2014) omtaler sertifisering av arkiver for å vise deres evne til å bevare og sikre digitalt materiale.
- c. Opportunities in Data Exchange gir innspill til hvordan ulike instanser kan støtte arbeid med forskningsdata gjennom forskningsdatas livssyklus.

Dokumentene resulterte i en liste med kartleggingskriterier med til dels detaljerte spørsmål omkring funksjonalitet i arkivet.

Kartleggingskriteriene utgjorde grunnlaget for innhenting av informasjon om arkivene. Det neste steget ble å gjøre en gjennomgang av nettsidene til arkivene og åpent, tilgjengelige dokumenter omkring arkivet. I løpet av gjennomføringsperioden ble NorStore Research Training Seminar annonsert og avholdt. Seminaret presenterte blant annet tilbud om lagring av forskningsdata fra NSD og NorStore, samt Forskningsrådets retningslinjer for tilgjengeliggjøring av forskningsdata. Tilstedeværelse på seminaret 7. januar var med på å gi utfyllende informasjon og sette grunnlaget for intervju.

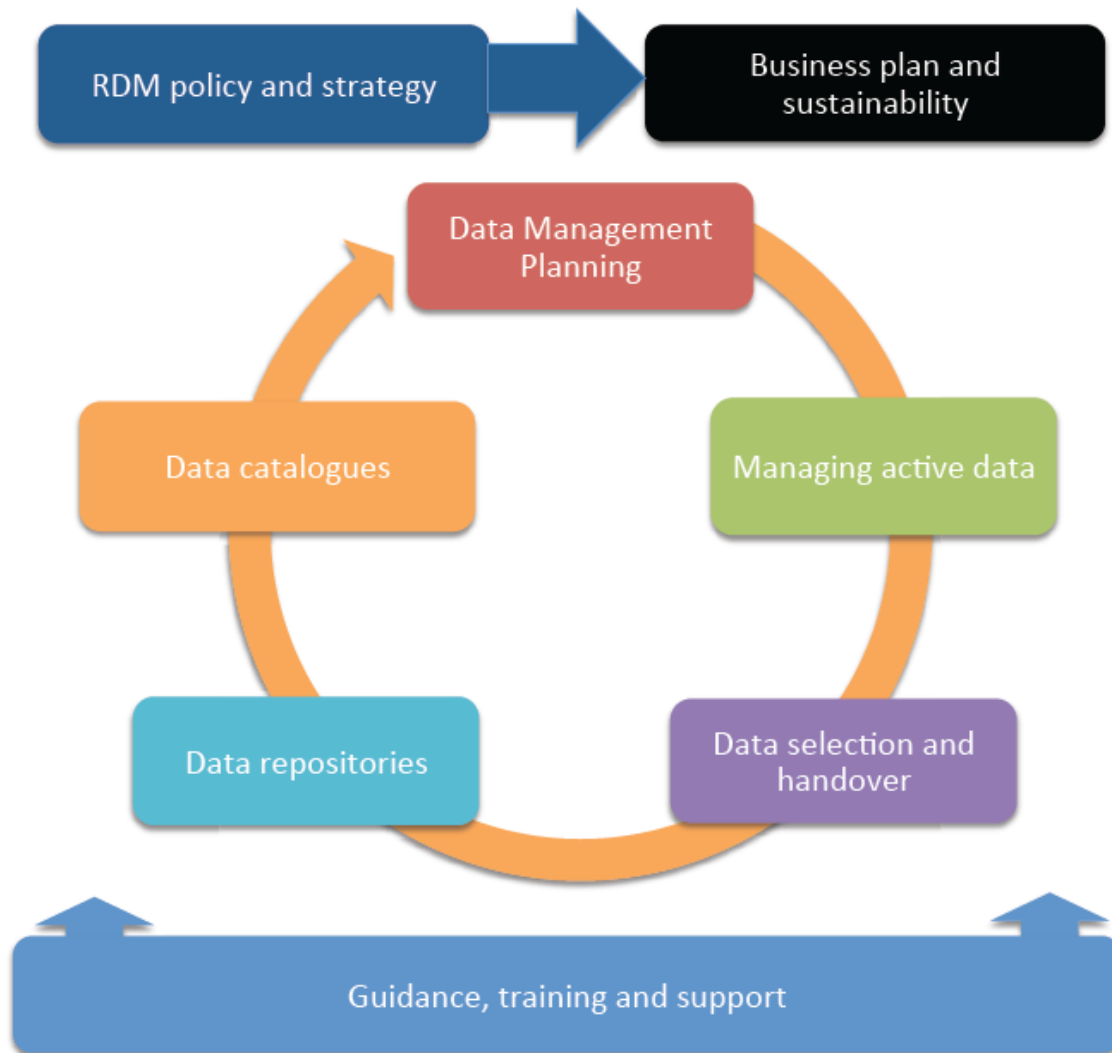
Tilslutt ble det gjort et personlig, semistrukturert intervju med representanter fra arkivene. Målet med intervjuet var å fange opp elementer fra kartleggingskriteriene som ikke kunne svares på gjennom nettsider eller dokumenter. Intervjuene ble fortatt februar 2015. Rammene for intervjuet ble satt til ca. 1 times varighet, og tidsrammen ble holdt. Undersøkelsen ble meldt til NSD personvernombudet for forskning.

I det videre vil det sees på lagring og bevaring av forskningsdata i form av datas livssyklus. Deretter vil det gjøres et utvalg av arkiver for forskningsdata som eksisterer i Norge. Arkivene vil vurderes etter kartleggingskriterier som er fremkommet gjennom en litteraturgjennomgang. Avslutningsvis vil det gis en beskrivelse av arkivenes tilbud med oppsummering av deres likheter og forskjeller.

Lagring og bevaring av forskningsdata

Lagring av forskningsdata er kun en del av forskningsprosessen. Det er en fase hvor data allerede er bearbeidet, og data er ofte anonymisert. Digital Curation Center viser at forskningsdata gjennomgår 5 stadier, omtalt som datas livssyklus (Jones et al., 2013):

Components of research data management support services



Figur 1. Modell over forskningsdatas 5 faser (Jones et al., 2013, s. 5.)

1. Forskerens planlegging og etablering av en datahåndteringsplan (data management plan) Dette skjer før forskningsprosjektet igangsettes.
2. Forskerens håndtering av data under forskningsprosessen (managing active data)
3. Utvelgelsen av data som skal tas vare på videre (data selection and handover). Foretas rett etter forskningsprosjektets avslutning.
4. Lagring av data i et dataarkiv (data repositories) som gjøres etter forskningsprosjektet og tilrettelegges for langtidslagring (data curation).

5. Deling av data som er lagret gjennom datakataloger (data catalogues) for at data skal kunne finnes tilbake til og kunne gjenbrukes. Jones et al. (2013) peker på at ved siden av dette er det behov for støtte og undervisning gjennom hele forløpet med datas livssyklus.

I dag har det blitt opprettet flere ulike typer arkiver, hvor forskeren kan velge å lagre data. Dette kan være arkiver som, først og fremst, er forbeholdt ansatte ved en bestemt forskningsinstitusjon, dvs. institusjonelle arkiver. Institusjonelle arkiv er vanligvis plassert organisatorisk i tilknytning til et bibliotek eller IT-avdelingen. Andre arkivformer kan være nasjonale datalagringsarkiver eller internasjonale arkiv. Arkivene kan være knyttet til ett eller flere faglige forskningsområder, eller til anvendelsen av bestemte forskningsmetoder. Det finnes også arkiver i tilknytning til forlag for vitenskapelig publisering, hvor det kan være krav om at lagring av data skal følge forskningspublikasjonen.

Noen forskningsfinansiører setter krav om bruk av bestemte arkiver, men de fleste har en henvisning til at det skal arkiveres data. Det kan være generelle anbefalinger omkring funksjonaliteten til arkivet når digitale forskningsdata skal lagres og eventuelt deles, som ved retningslinjene til EUs forskningsprosjekter finansiert gjennom Horizon2020 (European Commission, 2013).

Forskningsrådets policy for tilgjengeliggjøring av forskningsdata viser til at data skal være åpent tilgjengelig, når data er helt eller delvis finansiert med offentlige midler. Data kan unndras tilgjengeliggjøring på grunnlag av sikkerhet, personvern, juridiske eller kommersielle forhold. Andre forhold kan komme i betraktning, men da må det ha store konsekvenser av økonomisk eller praktisk art (Norges forskningsråd, 2014b).

Norske aktører for lagring av forskningsdata

Forskningsdata vil her bli definert slik som omtalt i retningslinjene til Forskningsrådet. Forskningsdata defineres som registreringer, nedtegnelser eller rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår i forskningsprosjektet. Kildedata, slik som kliniske data, værddata eller tekstkorpus, faller utenfor definisjonen. Grunnen er at data er innsamlet eller bearbeidet av andre (Norges forskningsråd, 2014b).

I denne kartleggingen er det av interesse å se på de største norske aktørene innenfor lagring av forskningsdata. Fokuset er nasjonale tjenester, hvor målet er å lagre og eventuelt dele data som er utarbeidet av den enkelte forsker eller forskningsgruppe, i tråd med retningslinjene til Forskningsrådet.

Ved at tilbudet er knyttet til nasjonale løsninger vil dette gjenspeile tilbudet som gir lagringsmuligheter for flest mulig forskere.

I Norge satses det på å bygge opp en infrastruktur for å tilrettelegge for forskning. Dette er blant annet beskrevet i norsk Veikart for forskningsinfrastruktur (Norges forskningsråd, 2014a)¹. Flere av prosjektene som er støttet omhandler forskningsdata og gjenbruk av data. Imidlertid er de fleste

¹ Populærvitenskapelig fremstilling i Avner (2014).

prosjektene begrenset til noen typer dataformater og et avgrenset fagfelt. Dette betyr at dataarkiveringstjenester tilbys av den enkelte utdanningsinstitusjon eller som er avgrenset til et bestemt fagmiljø, eks. Havforskningsinstituttets forskningsdata² faller utenfor fokuset for undersøkelsen.

Flere aktører i Norge tilbyr forskere muligheter for å bruke forskningsdata i egne studier. Data bearbeidet og innsamlet av andre kan finnes i en rekke norske arkivtjenester. Data brukt til forskning kan foreligge i registre som inneholder data samlet inn over lengre tid, gjerne statistiske opplysninger fra offentlige myndigheter eller medisinske helseopplysninger. Eksempler på slike register i Norge kan være Kreftregisteret, eller data tilgjengelig via Statistisk sentralbyrå. Beskyttelse av personvern er ofte årsaken til at denne typen data har egne lovmessige reguleringer og forskrifter. Basert på Forskningsrådets definisjon, så vil registerdata falle utenfor denne undersøkelsen. Nasjonalbiblioteket og Riksarkivet tilbyr nasjonale tjenester for lagring av dokumenter og offentlig informasjon som kan anvendes som et forskningsgrunnlag, men vil i dette tilfellet falle utenom kravet om lagring av digitale data.

Fokuset er på tjenester for ferdig bearbeidede data i tråd med Forskningsrådets kriterier. Andre nasjonale tjenester slik som Tjenester for sensitive data (TSD) som tilbys av NorStore og USIT ved UiO faller utenfor denne studien, da fokuset er på prosessen hvor forskeren bearbeider og analyserer materiale.

Norske instanser for lagring og avlevering av forskningsdata med nasjonalt tilbud er et begrenset utvalg. I denne studien er det fokus på to tjenester. Arkivløsningene for forskningsdata leveres fra Norsk samfunnsvitenskapelige datatjeneste [NSD] og fra NorStore.

Hvilken funksjonalitet bør arkivet ha?

Internasjonalt er det laget en rapport som viser ønsker for en forskningsdatainfrastruktur, «Strategy for a European Data Infrastructure». Denne rapporten er utarbeidet av Partnership for Advanced Data in Europe (PARADE). PARADE består av flere brukersamfunn og nasjonale partnere som arbeider for å forbedre datainfrastrukturen og fremme europeisk samarbeid. Målet er å skape en effektiv tjeneste som ser på data management behovet hos flere forskningssamfunn. Sammenslutningen har utarbeidet PARADE White Paper (2009) med en rekke tekniske karakteristika som de mener vitenskapelig e-infrastruktur bør inkludere. Karakteristika ved infrastrukturen fremhever blant annet:

- Åpen lagringsmulighet, slik at brukeren kan lagre data lett
- Utført slik at en gruppe forskere kan dra nytte av de lagrede data
- Permanente lenker som gir arkivene muligheter til å følge data tilbake til deres opprinnelse og beskriver karakteristika ved data
- Metadataregistrering slik at data effektivt kan administreres, brukes og forstås
- Interoperabilitet mellom systemer, både format og utveksling av informasjon i registreringsfelt

² Havforskningsinstituttets forskningsdata: <http://www.imr.no/forskning/forskningsdata/nn-no>

- Regulering av tilgang og lisenser for data
- Lagringsmåter som sikrer at data kan anvendes over et bestemt tidsrom
- Høy pålitelighet slik at forskeren kan stole på tilgjengeligheten av data.

I den overnevnte rapporten legges det stor vekt på den tekniske strukturen for avlevering, lagring og deling av forskningsdata. Flere nyere rapporter og policy initiativer legger vekt på muligheter for gjenbruk av data. «Riding the wave how Europe can gain from the rising wave of scientific data» (Wood et al., 2010) tar for seg forventede fordeler med lagring og deling av data. Denne rapporten følges opp av «A surfboard for riding the wave» (Van der Graaf & Waaijers, 2011) som ser på praktiske aspekter for å kunne gjennomføre deling av forskningsdata. En nyere rapport om «Integration of Data and Publication: Opportunities in Data Exchange» tar i større grad for seg koblinger mellom data og publikasjoner (Reilly et al., 2011). Opportunities in Data Exchange-rapporten henviser til betydningen av referanser for datasett med permanente lenker. Å bli kreditert for sine data er viktig for forskeren. Informasjonselementer som inngår i referanser basert på standarder for siteringspraksis for datasett er det viktig for arkivene å inkludere.

De nevnte rapportene beskriver samlet en rekke funksjonalitet som kan anvendes i praksis på infrastruktur i arkiver. I det videre gis noen eksempler på hva funksjonaliteten kan innebære i praksis.

Muligheter til å følge data tilbake til deres opprinnelse og beskrivelse karakteristika ved data gjennom permanente lenker er et eksempel. Permanente lenker kan anvendes i de standardiserte formatene som for eksempel Digital Object Identifier (DOI) eller Handle.

Ved registrering av forskningsdata i arkiver vektlegges bruk av standardiserte metadata. Standardiserte metadata kan komme til uttrykk gjennom bruk av Dublin Core, ISO19115 for geografisk informasjon, Data Document Initiative (DDI), Metadata Encoding and Transmission standard (METS), General international standard archival description (ISAD (G)), Norsk arkivstandard (NOARK) eller Common European Research Information Format (CERIF). Standardisering av metadata er viktig for å kunne sikre semantisk interoperabilitet, slik at deling av metadata kan foretas, for eksempel gjennom Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

Registrering av tilgang og lisensiering av forskningsdata er viktig for å kunne definere muligheter for gjenbruk av data. Dette kan reguleres gjennom lisensieringsbetingelser satt opp i Norwegian License for Open Data (NLOD), Creative Commons eller Open Data Commons. Det finnes ytterligere flere standarder enn de som nevnes her, og temaområdene er flere slik som krav om sikkerhetsløsninger i arkivet og dokumentasjon omkring arkivets bruk.

Det kan være vanskelig å holde oversikt over alle de forskjellige kriteriene og funksjonaliteten. Data Seal of Approval (Dillo & De Leeuw, 2014) er en sammenslutning som gir sertifisering til arkiver for å vise at de oppfyller en retningslinje for sikkerhet og bevaring av digitalt materiale. Data Seal of Approval er utarbeidet på grunnlag av retningslinjer som tidligere er presentert av en rekke arkivinstanser, slik som Nederlands nasjonalbibliotek, UK Data Archive, Dataarchive for Networked Services (DANS). Dataarkiver kan selv gjennomføre en sertifisering via nett, hvor man fyller ut et skjema med opplysninger om rutiner og prosesser i arkivet. Opplysningenes gjennomgå av en fagkomite som tildeler et sertifikat. Sertifikatet henviser til at arkivet holder en viss kvalitetsmessig standard. Sertifiseringen er kun gyldig i et avgrenset tidsområde, hvor resertifisering må gjøres etter et par år.

Data Seal of Approval vurderer arkiver med hensyn hvordan betingelser knyttet til områder som angår den som skal avgi data, arkivets funksjonalitet og gjenbruk av data. Bruk av arkivet for forskeren som har utført undersøkelsen og vil avgi data, setter betingelser knyttet til kvalitet og dokumentasjon av data. Arkivets funksjonalitet er for eksempel knyttet til dokumentasjon av rutiner og prosedyrer for administrasjon og bevaring av data i arkivet i tråd med datas livssyklus. Arkivet skal sikre tilgjengelighet for data ovenfor personer som avgir data og sørge for at andre kan gjenfinne og gjenbruke data. Arkivet skal følge internasjonalt anerkjente arkivstandarder. Ved gjenbruk av data skal arkivet følge det som er aksepterte normer/regler for gjenbruk og utveksling av data innen sektoren og forplikte seg til å følge lisensieringsbetingelsene som er satt på data.

Retningslinjene ved søknad til forskningsmidler fra EU Horizon 2020 fremhever at forskningsprosjektene skal ha en datahåndteringsplan. I vedleggene, annex 1 og annex 2, gis det forslag til elementer som kan inngå i en datahåndteringsplan. Noen aspekter som forskeren må ta stilling til er følgende: Er det kobling mellom datasett til øvrige vitenskapelige publikasjoner? Er data lagret i sertifiserte arkiver for langtidsoppbevaring og bevaring? Hva skal til for at andre skal ha tilgang til data? Vil data ligge åpent tilgjengelig, eller er det bare enkelte brukergrupper som skal ha tilgang. Hvilket arkiv skal data lagres i?

Forskningsrådets retningslinjer følger mange av de samme linjene som EU Horizon 2020, hvor det vektlegges at data skal lagres og arkiveres på en sikker måte. Semantisk interoperable metadata bør gjøre andre i stand til å søke etter og ta i bruk dataene, følge internasjonale standarder og gi en beskrivelse av datakvaliteten. Metadata bør gjøres tilgjengelig uten kostnad og publiseres slik at de kan høstes maskinelt og brukes i søk etter forskningsdata.

Samlet sett har gjennomgang av PARADE White Paper (PARADE, 2009), Opportunities in Data Exchange (Reilly et al., 2011) og Data Seal of Approval (Dillo & De Leeuw, 2014) og forskningsfinansierens (European Commission, 2013; Norges forskningsråd, 2014b) krav resultert i en liste med momenter som kan belyses ved vurdering av arkiver som tilbyr lagring og bevaring for ferdig bearbejdede forskningsdata. På de følgende sider vises kartleggingskriteriene slik de fremstod etter en gjennomgang av litteraturen. Kriteriene ble brukt i forbindelse med vurdering av Norsk samfunnsvitenskapelige datatjeneste – arkivet og Norstore Research Data Archive.

Kartleggingskriterier/ Intervjuguide

Innledning/velkommen

- Takke for at fikk komme og at vedkommende ville delta
- Opplyse om prosjektets formål
- Opplyse om generelle temaer som ønskes å ta opp
- Opplyse om personvern og samtykke
- Opplyse om lydopptak
- Opplyse om anonymisering
- Opplyse om samtykkeerklæring
- Har vedkommende noen spørsmål før vi starter intervjuet?

Informasjon om arkivet

Hva er nettadressen til arkivet?

Bakgrunnsmateriale/dokumentasjon: (referanser)

A. Beskrivelse av arkivet

1. Overordnet mål:
Hva er hensikten med arkivet?
2. Organisasjon/ Eierstruktur:
 - a. Hvem er det som står bak? Hvor mye penger og personalressurser som investeres?
 - b. Hvilken type arkiv er dette?
 - c. Hvor lenge har tjenesten eksistert?

B. Innflytelse

3. *Hva synes de om politiske aspekter/ tiltak for å øke lagring av data? Hvilke tanker har de omkring andre praktiske aspekter – intensiver, o.l.? (Hva betyr politiske initiativer /myndighetenes retningslinjer for deres virksomhet?)*
 - a. Har arkivet internasjonale eller andre nasjonale samarbeidspartnere?
 - b. Hvilke organisasjoner/ (politiske) finner dere som retningsgivende for sine tjenester innenfor feltet datalagring og deling?
4. Arbeider dere med å være en innflytelse på beslutningstakere som gir (politiske) retningslinjer/ føringer for drift? (KD, NFR, UHR)

C. Formidling av arkivets formål/virksomhet

5. *På hvilken måte formidler arkivet informasjon om egen virksomhet?*
 - a. *Informasjon om arkivet (explisit mission – informasjonsbrev?)*
 - b. *Strategi for arkivet. Ligger policy/ informasjon om planlagte endringer tilgjengelig?*
 - c. *Hvordan argumenterer arkivet for sin egen eksistens, arkivets impact eller innflytelse på research impact)*
 - d. *Er det noen publikasjoner som beskriver arkivet og funksjonaliteten?*
6. Dokumentasjon virksomheten til arkivet: Dokumenteres prosessen og prosedyrer/rutiner for administrasjon av lagrede data? Er dette koblet til arbeidsflyt i *data life cycl*us?

D. Om datalagring

7. Tilgang til datalagring:

- a. Hvem kan lagre data der? *Er det noen fagområder som anvender arkivet mer enn andre? Hvilke fagområder lagrer mest informasjon i arkivet?*
- b. Hvordan får forskerne tilgang til systemet?
Hva må til for at forskeren skal kunne laste opp data? Pålogging?
- c. Hvordan er tilgangen – online? Kan forskerne bruke plattformen til å bearbeide data?
Er det tilgang til en eller flere samtidige brukere?

8. Innhold:

- a. Hva lagres der?

9. Type data i arkivet:

- a. Hvilken type data er det mulig å lagre i arkivet?
- b. Er det noen begrensninger? Størrelse på datasett.
- c. I hvor stor grad er data i arkivet bearbeidet? For eks. Primærdata, sekundærdata, registerdata, ulike typer studiedesign?
- d. Forskningsetisk:
Er det koblinger til forskningsetiske godkjenninger med hensyn til personvern for informanter og studieobjekter?
Er det koblinger til finansieringsinstanser? Prosjektnummer?

10. Tekniske system: Hvilket teknisk system ligger til grunn?

Er systemet egenutviklet? Eller består det av en eller flere andre verktøy tilpasset til formålet?

11. Metadata:

- a. Hvordan beskrives data?
- b. Emnebeskrivelser følger de bestemte etter standarder?
- c. Kan metadata deles? (for eksempel høstes til eksterne digitale arkiver eller kataloger)

12. Sikkerhet: Hvilken sikkerhet har data?

13. Kontrakter/avtalerettslige aspekter: Hvordan reguleres avtaler omkring lagring? Er det egne kontrakter? Er dette en form for standardkontrakt, hvor ofte revideres kontraktbetingelsene?

14. Tilgang til data for andre /Opphavsrett:

- a. Hvem beholder opphavsrett til data lagret i arkivet?
- b. Kan datasett lastes ned?
- c. Hvem kan gjøre dette?
- d. Brukes det lisenser/ hvilke avtaler brukes for å regulere dette? I så fall hvilke er dette?

15. Sitering av data

- a. Er det oppgitt i basen hvordan brukere skal sitere data i egne publikasjoner?
 - b. Er det mulig å eksportere referanse informasjonen?
 - c. Tydelig markeringer av hvilken versjon av data brukerne henter ut av arkivet?
16. Permanent lenking: Tilbys permanente lenker? Til hele eller til deler av datasettet?
17. Statistikkinnsamling omkring bruk av data: Hvordan følger dere utviklingen av arkivet over tid? (Gjøres det vurdering av databasens relevans/ målinger over bruk? ((bruksdata, hvilke kilder brukerne kommer fra, hvilke nettsteder som henviser og hvor lang tid brukerne er))
18. Arkivering over tid:
- a. Er arkivet tilpasset langtidslagring? Hvor lenge kan data lagres der?
 - b. Hvordan bevares data over tid? Gjøres det oppdateringer på datasett, slik at det tilpasses ny programvare? Eller oppbevares eldre programvareversjoner? Legger arkivet vekt på lagring av programvare eller oppdateres innholdet i tråd med programvareoppdateringer?
 - c. Hvem har tilgang til data der? under hvilke betingelser?
19. Formidling av innhold om data i arkivet:
- a. Er det lagt inn automatikk for å holde brukere/andre forskere oppdatert med nytt innhold i arkivet? (RSS-feed, automatisk twitrning, eller andre verktøy for deling i sosiale medier? Vet du om arkivet er på liste over noen godkjente arkiver hos datatidsskrift? Eller har vært brukt som arkivkilde i andre tidsskriftartikler?)
 - b. Er arkivet indeksert i oversikter hvor forskere kan finne mulige kilder til data i deres egen forskning?
20. Tilvekst av data i arkivet:
- a. Er dere fornøyd med oppslutningen omkring arkivet blant forskere? På hvilken måte formidler dere tilbudet om lagring til forskerne (markedsføring)
 - b. Hvilket tilbud har dere til forskere som ønsker å benytte arkivet? Veiledning (en-til-en / forskergrupper) kurs hos dere eller hos universitetene/høyskolene
21. Gjenbruk av data:
- a. Hvilken type informasjon gjøres tilgjengelig for brukere/forskere som ønsker å benytte seg av data som ligger i arkivet? Henvises det til hvordan man kan få detaljer omkring data, slik som dokumentasjon av variabler, kodebøker?
 - b. Gis det noen form for krav til dem som ønsker å gjenbruke data? Krav til å følge av forskningsetiske normer, sitering av data o.l.

Avslutning

- Takk for samtalen.
- Tilsending av rapport når den er klar.

Beskrivelser av norske arkiver

Resultatene beskriver infrastrukturen som arkivene tilbyr. Beskrivelsen av arkivene vil kort gi bakgrunnsinformasjon om arkivet og dets formål. Forskning foregår i en internasjonal, global kontekst hvor utveksling av informasjon er sentralt. Dette betyr at de organisasjoner som tilbyr tjenester til forskning må befinne seg i en global kontekst. Det er et mål om at data utveksles digitalt, gjennom felles standarder for datautveksling. Av denne grunn er det et behov for å vite hvilke internasjonale organisasjoner som er viktig for arkivene, eller hvor de selv har en aktiv rolle. Eierskap er viktig for å si noe om aktørene som står bak, og gir mulighet til å trekke slutninger omkring arkivets levedyktighet over tid.

Betingelser for bruk av arkivet og hvordan data mottas i arkivet er knyttet til funksjonalitet og krav knyttet til bruk av arkivet for forskeren som ønsker å avlevere data. Arkivets registreringspraksis for data gjenspeiler hvilken informasjon gjøres tilgjengelig gjennom arkivet.

Ved bruk av arkivene for lagring og tilgjengeliggjøring av forskningsdata er funksjonaliteten førende for gjenfinning og deling. Beskrivelser av dataavleveringspraksis og betingelser og registreringspraksis med beskrivelse av data setter føringer for i hvor stor grad data kan gjenfinnes og deles. Dette betyr også at hvilke praksis for deling som arkivet tilrettelegger for er førende for i hvor stor grad data utveksles og benyttes av andre forskere. Delingspraksis med andre arkivorganisasjoner og registeroversikter og kriterier for gjenbruk av data for andre forskere.

I presentasjonen av arkivene vil det ikke skilles mellom informasjon som er kommet frem gjennom nettsider, dokumentasjon, på seminar eller intervju. Mesteparten av de tekniske beskrivelsene og funksjonaliteten kan finnes gjennom nettsidene og arkivenes brukerveiledning, samt opplyst gjennom NorStore Research Data Training Seminar. Intervjuet vektla i større grad den videre utviklingen til arkivet og arkivets kommunikasjon med brukermiljøene.

I det videre vil det kort presenteres bakgrunnen og formålet med arkivet. Forskerens dataleveringspraksis og betingelser for bruk av arkivet. Arkivets registreringspraksis for data. Delingspraksis med hensyn til muligheter for gjenbruk av data i arkivet for andre forskere, og den videre utviklingen som er planlagt i tilknytning til arkivvirksomheten.

NorStore Research Data Archive

Om arkivet – bakgrunn og formål

[NorStore Research Data Archive](https://www.NorStore.no/services/archive)³ er et arkiv for offentlig finansiert, publisert forskning. Arkivet muliggjør lagring og deling av digitale data, og har muligheter for å ta imot store datamengder. Tilbudet kan benyttes av forskere i hele landet ved alle fagfelt. Informasjon om datasettene som er lagret i arkivet er offentlig tilgjengelig.

NorStore Research Data Archive ble lansert som et pilotprosjekt i mars 2014, og skal evalueres i 2015. NorStore Research Data Archive er et tiltak under Uninett Sigma², en datterorganisasjon av Uninett.

³ NorStore Research Data Archive: <https://www.NorStore.no/services/archive>

Uninett er en selvstendig enhet under Kunnskapsdepartementet. Arkivet støttes av gjennom midler i en konsortiumordning, hvor midlene kommer fra Forskningsrådet, UiO, UiB, NTNU og UiT. Fra 1. januar i 2015 er det en endring av organisasjons- og finansieringsstrukturen, slik at ca. 2/3 av all finansiering kommer fra de 4 universitetene i motsetning til tidligere hvor størstedelen av finansieringen kom gjennom midler fra Kunnskapsdepartementet.

Internasjonale organisasjoner

NorStore er med i flere internasjonale organisasjoner eller sammenslutninger. NorStore er en del av Nordic eInfrastructure Collaboration [NeIC]⁴. Organisasjonen har medlemmer fra alle de nordiske landene og består av tekniske eksperter fra tungregningssentre, «high performance computing» [HPC]. Målet er å bidra til utvikling av en e-Infrastruktur på områder som har felles nordisk interesse.

Et annet initiativ hvor NorStore deltar er sammenslutningen EUDAT⁵, et nettverk for samarbeidende dataarkiver, finansiert under EUs 7. rammeprogram. Målet er å støtte utviklingen av en felles datainfrastruktur og utvikle tjenester som er med å støtte opp under samarbeid, sikker lagring, gjenfinning og deling av data. NorStore Research Data Archive er med i videreføringen av EUDAT2020, fra mars 2015.

Dataavleveringspraksis og betingelser

De fleste brukerne av NorStore er forskere som utfører prosjekter innen tungregning. Dette betyr at de har behov for en arbeidsflate hvor det er mulig til å analysere og bearbeide data, før data arkiveres og eventuelt gjøres tilgjengelig for andre forskere og allmennheten. NorStore Project Area kan anvendes av en forskergruppe for felles tilgang til data, men ikke som samarbeids- eller bearbeidingsplattform. For å få tilgang til prosjektområdet, så må det søkes om at prosjektet og deltakerne godkjennes på forhånd. Kriteriene er at man må være medlem av et pågående prosjekt som er helt eller delvis finansiert av en offentlig, norsk finansieringsinstans. Da kan alle medlemmer av prosjektet få samtidig tilgang til data som inngår i eget prosjekt. Tjenesten kan brukes på alle typer forskningsdata som ikke lenger betraktes som sensitive. I dag har innholdet i arkivet en hovedvekt på forskning fra klima- og geovitenskap.

Praksis for prosjektområdet skiller seg fra rutiner for en forsker som kun ønsker å lagre data i NorStore Research Data Archive, og eventuelt dele data i etterkant av vitenskapelig publisering. Det er fremdeles krav om at forskningen skal være offentlig finansiert, og forskeren må tildele data en open access lisens som følger Norwegian License for Open Data eller Creative Commons V4.

For en forsker som ønsker å laste opp data til arkivet kreves brukernavn og passord. De fleste forskere kan bruke sitt vanlige brukernavn og passord, siden dette er knyttet til brukertilgangen som universiteter og høyskoler har gjennom autentiseringstjenesten FEIDE⁶, felles elektronisk identitet.

Datasett kan lastes opp via en nettløsning. Dersom datasettet er større enn 5 GB, så kan kommandobasert opplasting brukes. Filene blir låst i den versjonen som de lastes opp. Det er lagt en brukeravtale, «Terms & Conditions», som forskeren må bekrefte før han laster opp data.

⁴ Nordic eInfrastructure Collaboration: <http://neic.nordforsk.org/>

⁵ European Data Infrastructure: <http://www.eudat.eu/>

⁶ FEIDE: <https://www.feide.no/>

Brukeravtalen innebærer blant annet at forskeren er forpliktet til å registrere prosjektets ID til Forskningsrådet og om NorStore eller Nortur er brukt som forskningsinfrastruktur i Cristin.

Registeringspraksis – beskrivelse av data

Når en forsker har lastet opp data, så må data beskrives, og forskeren må eventuelt bidra med tilleggsmateriale som forklarer data. I NorStore må forskeren beskrive data, gi metadata, innen 3 måneder etter at de har lastet opp data. Metadata består av en del hvor man beskriver data uavhengig av fagområde, gjennom et skjema basert på Dublin Core med faste felter. Utenom beskrivende metadata, så er NorStore åpen for å legge til fagspesifikke metadata. NorStore har i dag lagt til metadata som følger en standard for geografisk informasjon. Metadata kan legges til basert på fagmiljøenes behov, men dette forutsetter enighet om metadatastandarder i fagmiljøene.

Ved slutten av registeringsprosessen har forskeren selv mulighet til å se hvordan data vil fremstå og ha muligheter til å gjøre modifikasjoner. Data kan kobles sammen med vitenskapelige publikasjoner. Når data først er lastet opp, så kan ikke disse endres eller slettes. NorStore garanterer langtidslagring, dvs. lagring i 10 år fremover. Arkivet er åpent for ustrukturerte data, dvs. data som vanskelig kan indekseres i en database.

Delingspraksis

Bruk av data for andre forskere

Alle som ønsker det kan få informasjon om innholdet i NorStore Research Data Archive via nettsidene. Katalogposter med beskrivelse av data og en permanent lenke til datasettet er åpent tilgjengelig. NorStore benytter permanente lenker via DOI med deres eget prefiks, doi:10.11582. Det finnes ingen norsk instans for DOI per dags dato. NorStore har inngått en avtale med den danske DOI instansen ved Danmarks Tekniske Universitet.

Dersom andre vil ha tilgang til lagrede forskningsdata er det regulert av forskerne som har lastet opp data gjennom lisensbetingelsene, dvs. Norwegian License for Open Data eller Creative Commons V4.

Per dags dato må personer som ønsker å gjenbruke data laste ned hele datasettet, man kan ikke velge ut deler av dette. Åpne tilgjengelige datasettet som er små datasett, mindre enn 5 GB, lastes ned via nettsidene. Store datasett gis det tilgang til via NorStore Project Area, hvor man må ha en konto (Feide). Har man ikke dette, så kan man få tilsendt en URL, med en begrenset varighet.

Deling av data med andre organisasjoner / arkiver

Metadata deles med andre organisasjoner som har mulighet til å høste data fra NorStore. OAI-PMH protokoll og Dublin Core metadata muliggjør denne prosessen. NorStore Research Data Archive er registrert i kataloger over tilbydere av forskningsdata, slik som Registry of Research Data Repositories⁷.

Videre utvikling av arkivet

NorStore vil i tiden fremover ha fokus på teknisk infrastruktur i arkivet, for å kunne gå fra pilotversjon av arkivet til en permanent driftstjeneste. Et prosjekt innen NeIC vil arbeide for å innføre

⁷ Registry of Research Data Repositories: www.re3data.org

infrastrukturen i EUDAT til en del nordiske forskningsmiljøer som en del av EUDAT2020. NorStore vil bidra inn med personalressurser for videreutviklingen av EUDAT. En annen oppgave fremover er å få sertifisering gjennom Data Seal of Approval for å vise den kvalitetsmessige standarden til arkivet.

Utvikling av informasjonstjenester til fagmiljøene er et annet aspekt som vil vektlegges i nær fremtid. Informasjonstjenester er viktig med tanke på krav til forskningsprosjekter omkring utvikling av en datahåndteringsplan når det søkes midler fra Forskningsrådet. En annen informasjonsaktivitet er faglig formidling av innholdet i arkivet. Faglig informasjon kan blant annet finnes i META magasin⁸, på sikt ønsker NorStore å inkludere artikler i nettavisen om norsk og internasjonal forskning, forskning.no.

Norsk Samfunnsvitenskapelig Datatjeneste – Arkivet

Om arkivet – bakgrunn og hensikt/formål

NSD ble etablert i 1971. NSD er i dag organisert som et aksjeselskap eid av Kunnskapsdepartementet. NSD er en organisasjon utøver flere ulike tjenester til forskningssektoren. Tjenestene er knyttet til forskningsdata på en eller annen måte. NSD-arkivet⁹ er en av disse tjenestene. Andre tjenester som er kjent innen forskningssektoren er Database for høyere utdanning med statistikk om sektoren. Personvernombudet for forskning blir huset hos NSD, men er en tjeneste eiet av forsknings- og utdanningsinstitusjonene. På denne måten er Personvernombudet for forskning betraktet som uavhengig av NSD-arkivet og øvrige tjenester under NSD.

NSD er tilbyder av store mengder forskningsdata som er fremkommet gjennom undersøkelser og innrapportering av data til NSD. Data kan ha form av registerdata eller enkelt undersøkelser.

Tjenesten NSD-arkivet har fokus individdata som er fremkommet gjennom spørreundersøkelser utført av forskere i Norge. Kategorien norske spørreundersøkelser inkluderer statistiske undersøkelser produsert av forskningsinstitutter som Statistisk sentralbyrå og Transportøkonomisk institutt. Tilvekst av data til arkivet består for en stor del av de data som kommer fra prosjekt finansiert av Norges forskningsråd. Forskere som mottar støtte fra Norges forskningsråd til prosjekter innenfor samfunnsvitenskap, humaniora, medisin og helse, miljø og utviklingsforskning har en kontraktfestet plikt til å lagre en kopi av deres data hos NSD for gjenbruk til forskningsformål. Forskning som er foretatt i Norge behøver ikke være offentlig finansiert, slik at egenfinansierte studier som for eksempel masterstudents arbeider kan arkiveres.

Internasjonale organisasjoner

NSD er medlem av en rekke internasjonale organisasjoner. Dette er organisasjoner som arbeider for å bedre funksjonalitet i arkiver og mellom arkiver, slik som [Council of European Social Science Data Archives \[CESSDA\]](#), [Data Documentation Initiative \(DDI Alliance\)](#), [International Federation of Data Organisations for the Social Sciences \(IFDO\)](#).

⁸ META magazine: <https://www.notur.no/meta-magazine>

⁹ NSD-arkivet: <http://www.nsd.uib.no/solr/nsu>

NSD er også deltaker i organisasjoner som har til hensikt å dele data om utførte undersøkelser. Undersøkelsene inneholder data som kan anvendes ved sammenlignende studier, for eksempel [Inter-university Consortium for Political and Social Research \(ICPSR\)](#), [Luxembourg Income Study \(LIS\)](#), [International Social Survey Programme \(ISSP\)](#).

Arkivet mottok i januar sertifisering gjennom Data Seal of Approval.

Dataavleveringspraksis og betingelser

De samfunnsvitenskapelige fagområdene er de største brukerne av NSD-arkivet. Andelen brukere fra andre fagområder, som fra helseforskningen er i vekst. NSD har siden 1981 hatt ansvar for arkivering av data fra prosjekter med finansiering fra Norges forskningsråd og arkiveringsplikten har vært innarbeidet i kontraktsvilkårene. Arkivering i arkivet er ikke betinget av finansiering eller de nevnte fagområdene, slik at forskere fra andre fagområder kan undersøke mulighetene for lagring.

Bruk og tilgjengelighet av data reguleres gjennom at forskeren fyller ut en arkiveringsavtale og sender over data og dokumentasjonsfiler på e-post (ev. post) til NSD for videre bearbeiding og lagring. Dersom filene er for store til å sendes over e-post, benyttes Uninetts «filesender» tjeneste.

Registeringspraksis – beskrivelse av data

NSD-arkivet har muligheter for å lagre data over lang sikt, dvs. med en tidsramme over ti år. Når forskeren har besluttet å arkivere data, sendt over data og dokumentasjon vil NSD-arkivet registrere metadata omkring datasettet. NSD registrerer mange detaljer omkring data. For å lette gjenbruk av dataene, dokumenterer NSD alle datasett i verktøyet Nesstar Publisher ned på variabelnivå; spørsmålsformuleringer blir registrert (som grunnlag for søkesystemet) og referanser til eventuelle publikasjoner blir lagt til.

Dokumentasjon av data tar utgangspunkt i arkiveringsskjema, spørreskjema og eventuelle rapporter/sammendrag som er mottatt fra forskeren. Forskeren har en mulighet til å se dokumentasjonen omkring data før den tilgjengeliggjøres på nett. Dette gir forskeren mulighet til å kvalitetssikre beskrivelser av data. Bakgrunnsdokumentasjon kan eventuelt korrigeres for bedre forståelse for andre som ønsker å gjenbruke data.

Metadata beskrives gjennom standarder som benyttes av Data Documentation Initiative [DDI]: <http://www.ddialliance.org/Specification/RDF>. Dublin Core inndeling av felt benyttes også.

NSD-arkivet har et system for å ivareta endringer av datasett, hvor det holdes oversikt over alle endringer av data i arkivet. Her skilles det mellom ulike varianter av et datasett (endringer av data) og ulike utgaver (endringer i dokumentasjon). Det originale datasettet slik det kom fra forsker/produsent vil alltid være tilgjengelig, selv om nye utgaver eller varianter skulle bli opprettet.

Forskeren vil beholde eiendomsretten til data, og ha mulighetene for å bestemme hvor tilgjengelig dataene skal være for gjenbruk. Avhengig av lisensiering av data vil informasjon og eventuell tilgang til ferdig bearbeidede data, finnes på arkivets nettsider. Data kan være åpent tilgjengelig for nedlastning fra nettsidene, ellers så lånes data ut basert på en låneavtale. Låneavtalen har reguleringer som innebærer at tilgang for alle kan gis, kun for ikke-kommersielt bruk, for til bruk for forskning eller kun etter avtale med forskeren.

NSD anbefaler forskere om å arkivere data så snart som mulig etter innsamling og tilbyr en embargoperiode før publisering, slik at data først er tilgjengelige etter prosjektperioden eller når forskeren er ferdig med å publisere materiale basert på datasettene.

Delingspraksis

Bruk av data for andre forskere

Lisensreguleringene som forskerne har satt på data er avgjørende for hvordan andre forskere har anledning til å bruke data. En del brukere kan få tilgang til variabler i datasett over nett, og kan velge ut variabler for eget bruk. Hvis avtalereguleringer er mer restriktiv, så kan variabler bestilles, hvor det inngår en avtale for bruk, taushetserklæring og eventuelt veiledererklæring. Brukere av data må tegne en avtale, hvor de blant annet er forpliktet til å oppgi hvilke publikasjoner og rapporter hvor datasett som brukeren har fått tilgang til via NSD er anvendt. Forskerne skal oppgi referanse til datasettet når de utgir egne publikasjoner. NSD har en ordning med permanente lenker, men følger per dags dato ikke den internasjonale standarden DOI. Data som skal utleveres blir tilrettelagt for gjenbruk. Dersom brukeren ønsker personsensitive data, så må det i tillegg inngås en databehandleravtale eller data må anonymiseres før utlevering.

Andre organisasjoner

Metadata deles med andre organisasjoner som har mulighet til å høste data gjennom en OAI-PMH protokoll. NSD-arkivet er registrert i kataloger over tilbydere av forskningsdata, slik som Registry of Research Data Repositories og Databib.

Videre utvikling av arkivet

NSD-arkivet arbeider mer langsiktig med opprettelse av en webportal med mulighet for online registrering og opplasting av data. Dette vil forenkle arbeidsprosessen til forskeren ved avlevering av data og forenkle interne arbeidskrevende registreringspraksis for arkivet.

Informasjonstjenester ut til fagmiljøene blir satt fokus på fremover, og kommunikasjon via sosiale medier vil legges mer vekt på. Det er ønskelig å inngå tettere samarbeid med forskningsinstitusjonene, spesielt med tanke på Forskningsrådets innføring av krav om datahåndteringsplan.

Likhet og ulikheter i tjenester for lagring og bevaring av forskningsdata

Forskningsdataarkiv	NorStore Research Data Archive	Norsk samfunnsvitenskapelige datatjeneste
Eier / Deltakende organisasjoner	Uninett Sigma ² (UiO, UiB, NTNU, UiT)	NSD (aksjeselskap under Kunnskapsdepartementet)
Datakilder	Offentlig finansiert forskning med undersøkelser foretatt av enkeltforskere/forskningsgrupper	Undersøkelser foretatt av SSB, TØI, osv. + enkeltforskere
Hovedfokus	Mest studier fra geovitenskap/klimadata	Mest samfunnsvitenskapelige undersøkelser
Avlevering av data	Lastes opp over nett	Sendes på epost
Pliktig avleveringsinstans	Nei	Ja, avtale med forskningsrådet om avlevering av data innenfor områdene
Avtaleregulering for bruk av arkivet / brukeravtale	Ja	Ja
Lisensieringsavtale for bruk av data som lagres	Ja	Ja
Regelmessig sikring av lagrede data	Ja	Ja
Versjonshåndtering av data	Nei, lagres som nye datasett	Ja
Permanente lenker	Ja, DOI	Ja, eget system
Metadatabeskrivelser	Ja, Dublin Core	Ja, Dublin Core
Deling av metadata	Ja, OAI-PHM	Ja, OAI-PMH
Emneordsindeksering	Geographic Information Systems Standardiserte emneordsindekser	Data Documentation Initiative
Langtidslagring	Ja, Garanterer 10 år	Ja – til evig tid
Sluttbrukeravtale	Ja	Ja

Tabell 1. Noen sentrale karakteristika ved NorStore og NSD-arkivet

Resultatene viser at det finnes en rekke likheter og ulikheter i de to arkivenes praksis.

Det finnes en rekke likhetstrekk mellom arkivtjenestene for forskningsdata som tilbys av NorStore og Norsk samfunnsvitenskapelige datatjeneste. Finansiering fra Kunnskapsdepartementet/ Forskningsrådet har vært avgjørende for etablering av arkivene og drift av arkivene. Internasjonale kontakter og føringer er viktig for oppbygging og utvikling av arbeidet med arkivene.

Arkivene har som mål å være et lagringssted for forskningsdata over lang tid. Arkivene utviser en åpenhet for å ta imot alle typer filformater. Dette er litt mer begrenset i praksis basert på format og størrelse av forskningsdata som ønskes lagret og bevart.

Arkivene er opptatt av at forskerens data skal lisensieres slik at regulering av bruk og eierskap ivaretas. Internasjonale standarder legges til grunn for utvikling av arkivene. Standarder anvendes blant annet ved registrering av metadata og utvikling av informasjon. Postene i arkivet inneholder informasjon som kan brukes for å lage referansehenvisninger til datasett med permanente lenker. Bruken av permanente lenker er forskjellige med hensyn til type identifikatorsom anvendes, mens i forskermiljø er ofte krav ved siteringspraksis DOI.

Til tross for mange likheter mellom arkivene i studiet, så eksisterer det en noen forskjeller mellom NorStore og Norsk samfunnsvitenskapelige datatjeneste.

Arkivene er etablert med mange tiårs forskjell. Dette betyr at den teknologiske utviklingen har hatt en utvikling som virker inn på tilbudet som de to arkivene tilbyr per dags dato. NSD-arkivet er bygget opp gjennom Nesstar som er programmert og utviklet i tilknytning til NSD. NorStore Research Data Archive er bygget opp av moduler som kan skiftes ut.

Registreringspraksis for NSD-arkivet innebærer i mye større grad flere ledd før datasettet er lagret og eventuelt klart for gjenbruk. Noe av årsaken til dette skyldes karakteristika ved data som skal lagres og bevares. Registreringsgraden med hensyn til elementer i et datasett er forskjellig, for eksempel registrerer NSD ned på variabelnivå, mens dette ikke er hensiktsmessig med data knyttet til NorStore. Dermed er de to tjenestene slik de framstår i dag «semantisk» inkompatible. Det er med andre ord ikke mulig å sammenstille data fra de to kjeldene for gjenfinning, uten betydelig merarbeid, eller vesentlig tap av semantisk interoperabilitet.

Måten de to arkivene forholder seg til versjoner av data er forskjellig. NorStore vil registrere alle versjoner av et datasett som en ny versjon i et eget datasett. NSD vil registrere versjoner av det samme datasettet, og har et system for å holde oversikt over dette. NSD holder oversikt over hvilke versjoner av et datasett som er lånt ut til de forskjellige forskningsmiljøene, dersom lisensen til et datasett regulerer tilgangen for gjenbruk på denne måten.

Filformater er et annet aspekt som skiller de to arkivene. NSD tar imot alle filtyper som kan leses gjennom den programvaren som de har tilgjengelig. I praksis innebærer dette at de mest brukte filformatene, slik som de mest vanlige filtypene knyttet til tekstdokumenter, regneark, statistiske data, databaser, billedfiler, video, lyd og geografisk informasjon. NSD legger vekt på bevaring og muligheter for gjenbruk av filene, hvor tilgang til programvare og filformaters lesbarhet. NorStore tar i mot de samme filtypene, men legger også inn standard filtyper fra fagmiljø som skiller seg fra de mest brukte. Data som finnes i arkivet i dag er for eksempel mye klimadata, som skiller seg fra de tradisjonelle filtypene. NorStore legger vekt på oppdatering av filformatene i tråd med endringer i utvikling. NorStore kan lagre store datamengder, mens den tekniske infrastrukturen til NSD-arkivet er mer begrenset i så måte.

Tilveksten av data i arkivene fungerer i dag forskjellig. NorStore baseres per dags dato på forskernes egen innmelding av behov for lagring av data. NSD har en langvarig arkiveringsavtale med Forskningsrådet.

Funksjonalitetene i arkivene er lignende, slik at dette er noe overlappende mellom de to arkivene. Forskjeller i formater og størrelse på data som arkivene kan avleveres og bevares over tid gjør arkivene forskjellige, slik at de utfyller hverandre.

Avrundning

Denne rapporten legger vekt på dataarkivenes tilbud som forskere i Norge kan benytte seg av. Rapporten har vært selektiv med hensyn til hvilke tilbud som inkluderes, slik at i praksis finnes det flere arkiver som forskerne kan benytte. Litteraturgjennomgangen har vist til elementer av betydning for forskeren ved lagring av bevaring av forskningsdata. Dette er belyst gjennom beskrivelsen av arkivene til NSD og NorStore, hvor funksjonaliteten i arkivene har vært i fokus.

Arkivene er avtaleregulerte med hensyn til forskerens bruk av arkivet for lagring og bevaring av forskningsdata. Muligheter for embargoperiode før publisering av data er av betydning ved avlevering av data. Forskeren har selv anledning til å velge mellom forskjellige lisensmodeller for tilgjengelighet og gjenbruk av data. Dokumentasjon av data som skal lagres er sentralt for mulighetene for gjenbruk av data. Utvikling av standarder for dokumentasjon for eget fagfelt vil sannsynligvis bli større fokus på innen forskningsmiljøenes etter hvert. Gode eksempler og opplæring vil være sentralt for en god datapraksis. Anerkjennelse for egne forskningsaktiviteter er et intensiv for lagring av data. Arkivene registrerer aktuelle opplysninger for sitering med permanente lenker til informasjonen.

Registreringspraksis i arkivene viser at metadata fra arkivene følger internasjonale standarder, og muliggjør høsting til internasjonale oversikter som inneholder datasett. Arkivene er indeksert i internasjonale tjenester som viser tilgjengelighet av forskningsdataarkiver, slik at forskeren får hjelp til formidling av sine data ut over landegrensene. NSD-arkivet og NorStore Research Data Archive arbeider med å følge internasjonale standarder for forskningsdataarkiv.

Informasjonstjenestene omkring arkivene som er nasjonalt tilgjengelig vil styrkes. Fremover vil fokus være på informasjonstjenester ut til forskerne. Informasjon formidlet over nettsider og sosiale medier vil formidle innhold i arkivet (tilvekst av nye datasett), og gi informasjon om eksisterende tjenester og tilbud.

Rapporten tar for seg lagring av data og bevaring av data over tid med mål om å kunne gjenbruke og gjenfinne data. Dette er et av stegene i datas livssyklus. Dette betyr at innsamling, bearbeiding og dokumentasjon av data er knyttet til den individuelle forskerens eller forskningsgruppens arbeidsflyt. Dette er en oppgave som er knyttet til den enkelte forskningsinstitusjon i dag. I en videre utvikling kan det diskuteres om dette bør i større grad kobles til nasjonale tjenester, for å sikre smidige overganger mellom de ulike fasene i datas livssyklus, og sikre muligheter for gjenbruk i siste instans. Forskningsrådets krav om datahåndteringsplan innebærer et større fokus på lagring og bevaring av data, hvor målet er å øke gjenbruk av forskning. Gjenbruk av forskning forutsetter god dokumentasjon av data i arkivet, slik at opplæringstiltak for datahåndteringsplaner og god praksis for dokumentasjon er blant arkivenes planer.

Endringer med hensyn til krav om lagring av forskningsdata, forskningsfinansiører og forskningsinstitusjoner kan føre utviklingen av arkivene i en annen retning fremover. I løpet av de nesten årene antas det at måten forskningen og forskningsprosessen foregår på vil skape endring i praksis, hvor man ser en utvidet bruk av nye forskningsmetoder og filformater. Dette ser man blant annet gjennom tjenester som er under utvikling per dags dato. Det utvikles en lagringstjeneste for audiovisuelt materiale, E-Video. Dette er et samarbeid mellom flere aktører i sektoren som Norsk

samfunnsvitenskapelig Norsk samfunnsvitenskapelige datatjeneste, Universitetet i Oslo /USIT, og NorStore/Uninett Sigma².

Dersom man tar høyde for utviklingen i infrastrukturen i Norge som er planlagt, vil de nasjonale tjenestene fra dataarkivene gi de største brukergruppene av forskerne et tilbud for bevaring av data ved bruk av de mest vanlige filformatene.

Forskningsinstitusjonene har blitt oppfordret til å lage sine egne retningslinjer for datahåndtering og eventuelt innføre krav om en datahåndteringsplan. Dette har aktualisert diskusjoner ved utdannings- og forskningsinstitutt omkring publiserte, digitale datas plassering i et dataarkiv som allerede eksisterer, eller muligheter for opprettelse av lokale arkiver (eventuelt i tilknytning til en overordnet nasjonal tjeneste). Som nevnt tidligere er en oversikt over arkivenes tilbud og funksjonalitet viktig for å kunne foreta en begynnende avklaring om rolle/arbeidsfordeling mellom forskningsinstitusjoner, fag- og forskningsbibliotekene og dataarkivene.

Kriteriene til funksjonalitet i dataarkiv rundt lagring og bevaring av forskningsdata vil være like uansett om tjenesten er lokal eller nasjonal, som en følge av krav fra forskningsfinansiører og internasjonale interesseorganisasjoner. Et arkiv må oppfylle en rekke internasjonale standarder for funksjonalitet og være knyttet til krav om en lengre driftsperiode for å ivareta kvalitet og sikring av data. Funksjonaliteten i arkivet er avgjørende for at tilbudet skal være til det beste for forskeren som produsent av data og forskeren som gjenbruker av data.

Etterord

Rapporten hadde som mål å se på eksisterende tilbud innen arkivering og bevaring av forskningsdata innen universitets- og høyskolesektoren. Rapporten har tatt for seg to tjenester for bevaring av digitale forskningsdata som eksisterer på nasjonalt nivå. Arkivene dekker mange av tjenestene som forskerne har behov for på en sikker måte.

Kartleggingen hadde som et overordnet mål å se på hvilke roller biblioteksektoren/BIBSYS kan ivareta, og hvordan BIRD, som et lokalt arkiv kan forholde seg til andre arkiver i sektoren. Dette for å finne grenseflater og eventuelt se på hvordan roller kan utfylle hverandre.

Dette fordrer til å se på mulige roller og aktiviteter som bibliotekene kan ha innen forskningsdata i høyere utdanningssektoren. Internasjonalt har flere organisasjoner sett på ulike aktiviteter og roller som bibliotekene kan ta (European Commission, 2013; Reilly et al., 2011; Thestrup et al., 2012; Van der Graaf & Waaijers, 2011). Rapportene og føringene kan oppsummeres i 4 aktiviteter.

1. Metadatastøtte for bedre gjenfinning av datasett
2. Sikre muligheter for gjenbruk av datasett over tid
3. Lage institusjonelle lagringsløsninger for datasett
4. Veilednings- og rådgivningstjenester

Disse aktivitetene kan i seg selv utgjøre en rolle for bibliotekene, eller flere aktivitetene kan forekomme i en kombinasjon.

Metadatastøtte til gjenfinning av datasett er en av aktivitetene som kan utgjøre roller for bibliotekene. Bibliotekarar har arbeidet mye med metadata i forbindelse med katalogisering. I arbeidet med forskningsdata er det en forutsetning for gjenfinning at gode metadatabeskrivelser legges på datasettene. I tillegg er det behov for permanente lenker som knytter sammen metadata om datasett (katalogposter) og publikasjoner om datasett (fagartikler eller datatidsskriftartikler). Det pekes på at Data Management Plans bør kunne kobles sammen med annen informasjon omkring selve lagringen av data og publikasjoner.

Bibliotekene kan bistå andre deler av organisasjonen/datalagringssenter med anbefalinger omkring metadata og tilhørende internasjonale standarder som benyttes for gjenfinning av datasett.

Bibliotekene kan:

- Støtte arbeidet med permanente lenker og siteringsstandarder
- Oppmuntre til utviklingen av felles metadatabeskrivelsesskjema og felles siteringspraksiser
- Fremme bruk av felles standarder og verktøy blant forskere

Å sikre mulighet for gjenbruk av datasett over tid, er en annen aktivitet som kan inngå som rolle for bibliotekene. Gjenbruk av data er et vanskelig aspekt ved datalagring. Over tid vil programvare anvendt for å produsere datasett endres. Dette kan gjøre gjenbruk av data vanskelig, dersom man ikke har tilgang på riktig programvareversjon. Er data tilgjengelig i dagens versjon av statistikkprogrammet SPSS 22, så betyr ikke det at om 10 år vil det være mulig å åpne filen og bruke

data. Det vil kreve at data er tilpasset nye versjoner av SPSS og det nye operativsystemet som brukes på din datamaskin, eller gamle systemer er tilgjengelig for avlesing av filer.

Det er en forutsetning for gjenbruk av forskningsdata at man har kjennskap til fagfeltets forståelse av datasett, og tilgang til programvaren som er brukt i forbindelse med datasettet, og kunnskap om bruk av programvaren.

Bibliotekene kan bistå med å tilrettelegge for gjenbruk av forskningsdata ved å:

- Arkivere og bevare datasett
- Arkivere programvare nødvendig for gjenbruk/etterprøvbarhet/bedre utnyttelse av data
- Formidle til forskerne informasjon om fagspesifikke lagringsordninger/arkiver (formater for lagring, dokumentasjonsopplysninger om fremgangsmåte for datainnsamling, lisenser)
- Å vise under hvilke betingelser datasett kan gjenbrukes (forutsetninger for at andre forskere skal kunne anvende datasett og oppbevare tilhørende programvare)

Utvikling av institusjonelle arkiver for lagring av datasett er en tredje gruppering av oppgaver som kan inngå som roller for bibliotekene.

Norske forskere har i dag muligheter til å lagre data i forskjellige typer arkiver. Dette kan være arkiver som er forbeholdt ansatte ved en bestemt forskningsinstitusjon, dvs. institusjonelle arkiver.

Institusjonelle arkiver er vanligvis plassert organisatorisk i tilknytning til et bibliotek eller IT-avdelingen. Andre arkivformer kan være nasjonale eller internasjonale datalagringsarkiver eller arkiver innen faglige forskningsområder eller knyttet til bestemte forskningsmetoder.

I Norge finnes det institusjonelle arkivløsninger ved de fleste universitetene og høyskolene. De drives av bibliotekene. Arkivene inneholder publiserte tidsskriftartikler, masteroppgaver, rapportserier og annet materiale. På en del fagfelt, hvor det ikke finnes så mange etablerte praksiser, kan et institusjonelt arkiv for forskningsdata avhjelpe situasjonen.

Bibliotekenes erfaringer med denne typen arkivløsninger er verdifullt med hensyn til opprettelser av arkiver for forskningsdata. Kjennskap til Discovery tjenester (samsøk), informasjons- og metadatastandarder regnes som viktig funksjonalitet i arkiver for forskningsdata. Når forskningsdata skal lagres vil formatene på data som anvendes i mange tilfeller være andre enn i dagens arkiver, slik som for eksempel bilder eller lyd. Gjenfinning vil ha andre standarder og behov enn dagens institusjonelle arkiv for publikasjoner. Det finnes egne tekniske systemer produsert for å være institusjonelle løsninger for forskningsdata, men en del av dagens systemer kan utvikles for å ta høyde for andre filformater og informasjon slik at de kan benyttes til forskningsdata.

Veilednings- og informasjonstjenester er den siste kategorien med roller som bibliotekene kan være med å ta i arbeidet med forskningsdata. Informasjon omkring hvordan forskeren kan gå frem, og hvordan hun kan gjøre sine data tilgjengelig blir viktigere. Bibliotekene kan ta en rolle med å bistå med informasjonstjenester om tilgjengeliggjøring av data, og forståelse av datasett som er lagret.

Hjelp til tilgjengeliggjøring av data

Mange forskere vet ikke hvilke tjenester som er tilgjengelig for datalagring eller kan ha problemer mellom å velge mellom forskjellige tjenester. Biblioteket blir sett på som en aktør som kan bistå med informasjon om god dataadministrasjon, fordelene med datadeling, mulighetene som er tilgjengelig

innen bestemte fagområder og hva som er de beste og mest pålitelige tjenestene. Bibliotek kan inkludere datamateriale i deres egne kataloger for lettere gjennomfinning av tilgjengelige datasett. På denne måten kan terskelen for datalagring senkes.

Bistå med tilgjengeliggjøring av informasjon omkring hvor datasett kan finnes

Informasjon omkring hvor datasett kan finnes kan integreres i bibliotekets tjenester, slik at bibliotekets discovery-tjenester høster informasjon. Eventuelt kan det utarbeides en protokollhøsting til en felles norsk plattform, etter prinsippet bak NORA – Norwegian Open Research Archive som i dag administreres av CRISTin.

Bistå med tilgjengeliggjøring av informasjon som fremmer forståelse av datasett

For å kunne fortolke data er det nødvendig med gode beskrivelser av data, beskrivelser kan forekomme som datapublikasjoner eller som lenking til publikasjoner som bruker data.

Bibliotekene kan:

- Lage oversikter over hvor datasett kan finnes (discovery services for datasets)
- Integrere datasett i øvrige gjenfinningstjenester (bibliotekataloger)
- Etablere og opprettholde kunnskapsbaser om data og deres kontekst
- Støtte krysskoblinger/crosslinks mellom publikasjoner og datasett
- Gi støttetjenester omkring informasjon omkring standardiseringer slik som opphavsrett, lisensieringer, kvalitet på lagringstjenester basert på sertifiseringer og søknadsprosedyrer for gjenbruk av datasett.
- Gi hjelp til forskerne for å forstå metadatabeskrivelser for datasett

Det er mange roller som bibliotekene kan velge å fokusere på eller bestemme seg for å bidra inn med i forhold til forskningsdata. Rapporten tar for seg lagring og bevaring av forskningsdata. Ved en forskningsinstitusjon kommer alle fasene hvor forskningsdata berøres til, fra de første tankene og søknad om forskningsmidler til utveksling av informasjon og analyser av data blant forskerne innen eller mellom forskningsinstitusjoner til lagring og bevaring av forskningsdata. Forskningsdata er et felt med mange aktører og interessepartnere ved utføring av behandling og bevaring av data, og temaet berører teknologiske, økonomiske og politiske aspekter. Det er for tidlig å si at fag- og forskningsbibliotek skal ha en rolle i arbeid med forskningsdata. For øyeblikket kan man se for seg en situasjon som vil variere fra institusjon til institusjon, valg med hensyn til strategi og satsing, og ikke minst kompetanse for å løse oppgavene som følger.

I løpet av denne prosjektperioden har forskningsbibliotek i Norge begynt å ta en rolle i tilknytning til teknisk tilrettelegging og arkiv for digitale forskningsdata. Bak oppstarten av tjenestene står drivkrefter av ivrige forskere som fungerer som foregangspersoner med ønske om å sikre og dele data. Tilbudet er nystartede initiativ og foreløpig av en liten skala.

Bibliotekene har bistått med å opprette lagringsløsninger knyttet til bestemte fagområder knyttet til bestemte fagområder i nær tilknytning til forskerens arbeidsplass. «Trolling¹⁰» et arkiv for lingvistisk

¹⁰ Trolling: <http://opendata.uit.no/dvn/dv/trolling>

forskning ved Universitetet i Tromsø og «TORD¹¹» for eksplosjonsforskning ved Høgskolen i Telemark er eksempler på dette. Fagpersonene har hatt en aktiv rolle i utarbeiding av arkivene og miljøene har dyktig IT-teknisk personale, i tillegg til bibliotekkompetanse. Det er derfor ikke gitt at alle fag- og forskningsmiljøer i Norge vil kunne bistå med lignende tjenester. Rapporten nevner blant annet at det er mange elementer som skal på plass for å sikre et godt tilbud for lagring av forskningsdata. Systemteknisk er det en rekke elementer som bør innfris, slik som tilstrekkelige metadata, mulighet for arbeidsflyt og informasjon med andre tekniske system, muligheter for lagring over tid og eventuelt versjonshåndtering. Sikkerhet er et viktig tema. Dette gjelder sikkerhet med hensyn til systemtekniske elementer og tilgangsregulering for brukere, og sikkerhet med hensyn til personvern gjennom anonymisering og lisensiering for regulering av tilgang til lagrede data.

De nasjonale arkivene har mål om å ta en mer aktiv rolle i forhold til produksjon av datahåndteringsplaner, formidle informasjon om data i sine arkiver og eventuelt bistå institusjonene å bistå til en mer effektiv produksjon av dokumentasjon.

Datahåndteringsplaner er nært knyttet til søknader om forskningsstøtte. De fleste forskningsinstitusjoner har etablert et tilbud i tilknytning til forskningsadministrasjon for å bistå faglige ansatte for å innhente eksterne finansieringsmidler, blant annet ved søknader.

Dokumentasjon av data er nært knyttet til fagfeltenes metodiske utforming og praktisk gjennomføring av analyser. Spredningen av ulike metodiske tilnærmingsmåter er stor mellom fagfelt, for eksempel humaniora og realfag, men kan være bred innen et fagfelt. Forskningsinstitusjonene vil ha behov for å kunne tilby sikre plattformer eller teknikker for utveksling av informasjon mellom forskere i prosessen med å analysere data. Dette er ikke en del av arkivenes tilbud, men vil håndteres på institusjonelt nivå med hensyn til valg av tjenester. IT-infrastrukturen, lokalt og nasjonalt, er en forutsetning for effektiv informasjonsutveksling innen et forskningsprosjekt, slik som brukeridentiteter, tilgangsreguleringer og fildeling får betydning for arbeidsflyt. Bruk av programvare til metodiske analyseverktøy kan være viktig for effektiv dokumentasjon av forskningsprosessen med hensyn til muligheter for å replisere undersøkelser og legge grunnlag for longitudinelle studier. Dokumentasjon vil kunne bidra til verdifull informasjon dersom forskerne velger å legge ut forskningsdata åpent.

Det er sannsynlig at forskningsinstitusjonene i tiden fremover kan ha et ønske om å ha en bedre oversikt over forskningsdata som produseres i egen organisasjon. Per dags dato kan informasjon registreres i CRISTin, men dette er en registrering som vil være frivillig for den enkelte forsker/forskningsgruppe.

Uansett, hvilken rolle bibliotek velger å engasjere seg i eller ikke, så er en av de største faktorene i arbeidet kunnskap og samarbeid. Det er behov for kunnskap omkring eksisterende tjenester, planlagte tjenester, og muligheter for integrering i eksisterende eller fremtidige systemer blant de ulike aktørene som er involvert i forskningsdata. Aktørene er både personer lokalt i en høyere utdanningsinstitusjon og på nasjonalt nivå, som arbeider i tilknytning til arkiv, bibliotek og IT-tjenester, eller som arbeidet med å utføre forskningsadministrasjon og forskning. Kunnskap om forskerens behov som bruker vil være avgjørende for tilbud i tilknytning til høyere utdanningssektoren.

¹¹ Tord: Telemark Open Research Data – lukket i utviklingsfasen

Bibliotekene har behov for kompetanse på feltet forskningsdata – både teknisk utførelse med «beste praksis» og kommunikasjon mellom systemer, juridiske implikasjoner omkring lisenser, lovverk og open access for data, og veilednings- og søketjenester for gjenfinning av forskningsdata. Norske bibliotek i høyere utdanning har tradisjonelt arbeidet tett sammen og bidratt med erfaringsutveksling i sektoren, uavhengig av de tekniske systemene som ligger til grunn for praksisutøvelse. Spesielt på feltet forskningsdata, som er nytt for sektoren, vil erfaringsutveksling være viktig for etablering av mulige samarbeidsrelasjoner.

Samarbeid er nødvendig med alle aktørene i sektoren som forskningsdata angår. Dette for å lage bedre systemer og rutiner for å kunne utnytte eksisterende infrastruktur og systemer i universitets- og høyskolesektoren. Samarbeid vil være avgjørende for å finne hvordan roller kan utfylle hverandre, spesielt med tanke på at fagmiljøer og datas bevaringsverdi er forskjellig. Utveksling av forskning foregår stort sett på et internasjonalt nivå, slik at et samarbeid og erfaringsutveksling for å bli deltaker i en internasjonal kontekst er førende. Kunnskap og samarbeid vil være satsingspunkter for alle aktørene i tilknytning til forskningsdata for å bistå forskerne som produsent og gjenbruker av data på en best mulig måte.

Referanser

- Avner, Terje. (2014, 30.04.2014). 50 prosjekter som skal gjøre norske forskere enda bedre. *Aftenposten*. Lastet ned 04.03, 2015, fra <http://www.aftenposten.no/fakta/innsikt/50-prosjekter-som-skal-gjore-norske-forskere-enda-bedre--7550233.html>
- Borgman, Christine L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge: MIT Press.
- Dillo, Ingrid, & De Leeuw, Lisa. (2014). *Data seal of approval: Certification for sustainable and trusted data repositories (2014-2015)* Lastet ned fra http://datasealofapproval.org/media/filer_public/2013/09/27/guidelines_2014-2015.pdf
- European Commission. (2013). Guidelines on Data Management in Horizon 2020: The EU Framework Programme for Research and Innovation. Lastet ned fra URL
- Jones, Sarah, Pryor, G., & Whyte, A. (2013). *How to develop RDM Services – a guide for HEIs* Lastet ned fra http://www.dcc.ac.uk/sites/default/files/documents/publications/How-to-develop-RDM-services_finalMay2013rev.pdf
- Kruse, Filip, & Thestrup, Jesper B. (2014). Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark. *LIBER Quarterly*, 23(4), 301-335. <https://liber.library.uu.nl/index.php/lq/article/view/9173/9922>
- Kuipers, Tom, Hoeven, Jeffrey van der, & Insight, Parse. (2009). *Insight into digital preservation of research output in Europe* Lastet ned fra http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Kunnskapsdepartementet. (2013). *Lange linjer - kunnskap gir muligheter*, hentet fra: <http://www.regjeringen.no/nb/dep/kd/dok/regpubl/stmeld/2012-2013/meld-st-18-20122013.html?id=716040> (Vol. 18(2012-2013)). Oslo: Departementenes servicesenter, Informasjonsforvaltning.
- Nielsen, Hans Jørn. (2012). Videnskabelig kommunikasjon og forskningsbiblioteker i en digital tid. I H. Høyrup, H. J. Nielsen & B. Hjørland (Red.), *Viden i spil: forskningsbibliotekernes funksjoner i forandring* (s. 67-92). Frederiksberg: Samfundslitteratur.
- Norges forskningsråd. (2014a, 05.05.2014). Samlet oversikt over veikartprosjekter. Lastet ned 04.03, 2015, fra http://www.forskningsradet.no/prognett-infrastruktur/Artikkel/Samlet_oversikt_over_veikartprosjekter/1253995493903
- Norges forskningsråd. (2014b). *Tilgjengeliggjøring av forskningsdata* Lastet ned fra http://www.forskningsradet.no/no/Nyheter/Forskningsdata_skal_deles/1254000298821/p1174467583739
- PARADE. (2009). *Strategy for a European Data Infrastructure* Kimmo Koski, Claudio Gheller, Stefan Heinzl, Alison Kennedy, Achim Streit & Peter Wittenburg (Red.), Lastet ned fra <http://www.csc.fi/english/pages/parade>
- Reilly, S. , Schaller, W. , Schrimpf, S. , Smit, E. , & Wilkinson, M. (2011). Integration of Data and Publication; ODE report: Opportunities in Data Exchange, hentet fra: http://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf.
- Thestrup, Jesper B., Kruse, Filip, Nondal, Lars, Dorch, Bertil F., Andersen, Mikkel, Blaabjerg, Niels Jørgen, . . . Knudsen, Ellen V. (2012). *Forvaltning af forskningsdata i Danmark: En undersøgelse af danske universiteters og forskningsråds praksisser, politik og visioner for lagring, langtidsbevaring, tilgjengeliggjørelse og deling af forskningsdata*. Hentet fra: <http://www.deff.dk/aktuelt/artikel/forvaltning-af-forskningsdata/>
- Van der Graaf, M. , & Waaijers, L. (2011). A Surfboard for Riding the Wave. Towards a four country action programme on research data. A Knowledge Exchange Report, hentet fra www.knowledge-exchange.info/surfboard.
- Wood, John, Andersson, Thomas, Bachem, Achim, Best, Christoph, Genova, Françoise, Lopez, Diego R., . . . Hudson, Richard L. (2010). *Riding the wave : How Europe can gain from the rising tide*

of scientific data. Hentet fra: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>: European Union.